The Messy World of Grey Literature in Cyber Security

Patricia A. Erwin
Institute for Information Infrastructure Protection. Dartmouth College, United States

Abstract
The foundation of my paper is based on four observations: 1) Research is messy; 2) Traditional collection development policies are structured documents aimed at assuring a level of quality in the collection, but also to satisfy the administrative need to justify the expense of providing resources to an academic or research community; 3) Grey literature doesn't fit the formal model of scholarly communication, therefore the quality is suspect and is not adequately addressed in most collection development policies; 4) The research process and grey literature share similar attributes. Libraries, by focusing on the formal products of research, miss the fertile, albeit more messy grey literature. I will use the I3P's focus on grey literature in cyber security as illustrative of how we need a broader definition of what constitutes the 'fruits' of research.

When we think of library collections we tend to think in terms of systems, order, and prescribed ways of tending to our collections. Our academic and research libraries are bound by the orderly world of academic departments, curriculum development, and the infrastructure of academia. Our collection development policies reflect that sense of order, clearly articulating the subject areas in which we actively collect, formats to be collected, and most importantly, the quality of the resources that will grace our shelves and gain a spot in our catalogs. What a tidy world we live in.

The truth is that research is messy. It is that intersection of the serendipity, randomness, and discovery that lends excitement to the research process. The chance merging of two seemingly unrelated concepts moves research into new areas of knowledge. These early findings and concepts do not appear in the standard scholarly communications vehicles, but rather in lab notebooks, concept papers, and technical reports, i.e. grey literature. There is a perception that grey literature is of less value than resources published through the more traditional and formal models of scholarly communications. In fact it has been noted that "scientific research is recognizable as such not because of the conditions under which it is performed but because of the way it is presented and published" (Pierce 1990, p. 55).

To better support research efforts, our collections must mirror that messiness of research. Traditionally, most collection development policies have not reflected the value of grey literature in the areas of computer science, and specifically cyber security. In support of my observations, I will report on a sampling of collection development policies from the I3P Consortium members' libraries. Our members represent academic research institutions, federal research labs, and not-for-profit research organizations. While much grey literature is collected internally, either by individual researchers or as part of an organizations institutional assets, the impetus for making this research widely available is mired in financial constraints, 'ownership' issues, and an underlying suspicion by some librarians that grey literature is not very quite as valuable and other resources that have moved through the publication process.

Overview of the I3P

The Institute for Information Infrastructure Protection (I3P) is a multi-organization consortium of academic institutions, federally funded research and development centers, and not-for-profit research organizations. The Consortium was founded in 2002 to bring experts together to identify and help mitigate threats aimed at the U.S. information infrastructure.

The I3P was founded in September 2001 to help meet a well-documented need for improved research and development (R&D) to protect the Nation's information infrastructure against catastrophic failures. The Institute's main role is to coordinate a national cyber security R&D program. Through the funding of research projects, the I3P works to identify and address critical problems in information infrastructure protection.

The information infrastructure consists of technologies and capabilities for gathering, handling, and sharing information that are accessible to, or commonly depended upon by, multiple organizations, whether within a single enterprise, a critical infrastructure sector such as banking and finance, the U.S. Government, the nation as a whole, or trans-nationally. The information infrastructure, taken as a whole, is not an engineered system. It is the result of the entrepreneurial efforts and the collective genius of the nation, working to improve efficiency and provide new opportunities for people and businesses. Security was not a major consideration at its inception, and security concerns today do not override market pressures for new uses of technology or innovation, in spite of frequent mention of hackers,